Figure1: Flow diagram of GeneDecipher
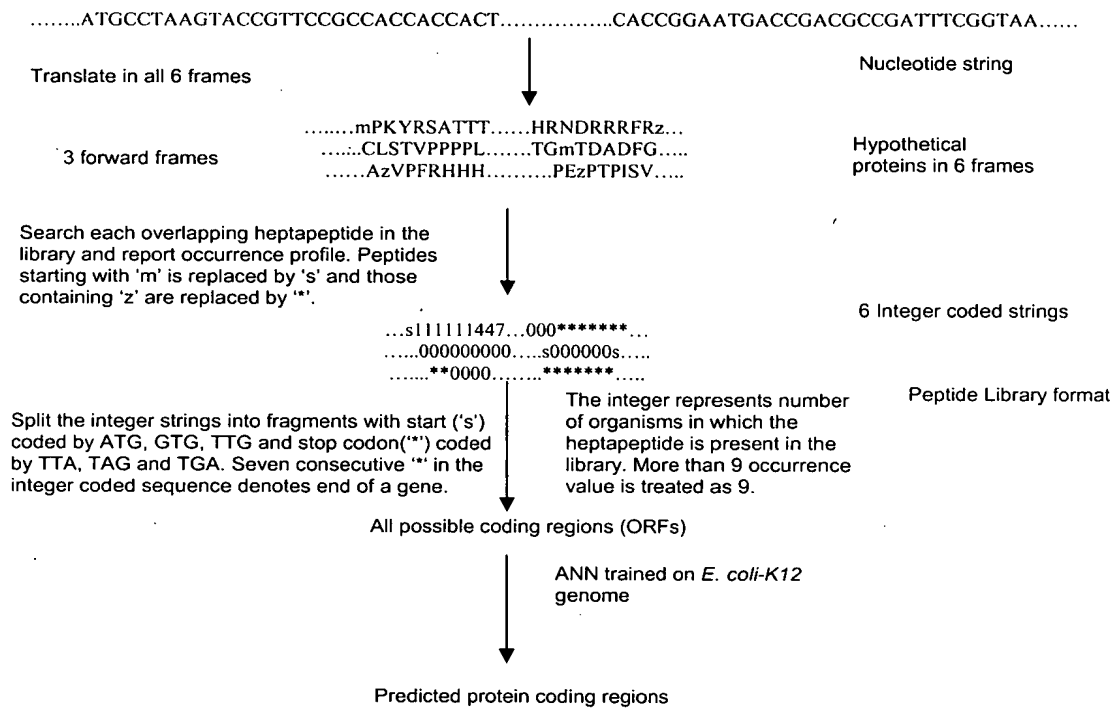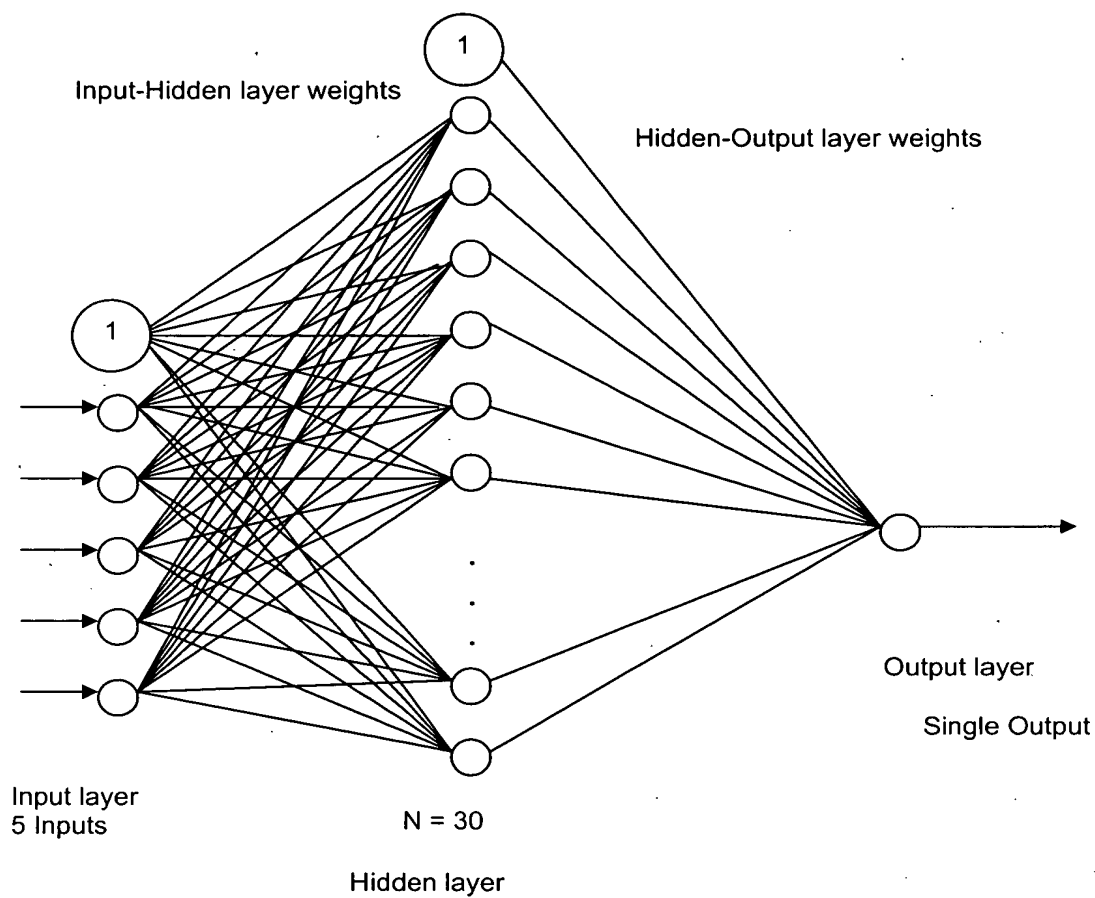
.........ATGCCTAAGTACCGTTCCGCCACCACCACT.................CACCGGAATGACCGACGCCGATTTCGGTAA......

Translate in all 6 frames

Nucleotide string

3 forward frames

........mPKYRSATTT......HRNDRRRFRz...
.......CLSTVPPPPL.......TGmTDADFG.....
......AzVPFRHHH..........PEzPTPISV.....

Hypothetical
proteins in 6 frames

Search each overlapping heptapeptide in the
library and report occurrence profile. Peptides
starting with 'm' is replaced by 's' and those
containing 'z' are replaced by '*'.

6 Integer coded strings

...s111111447...000*******...
......000000000.....s000000s.....
.......**0000........*******.....

Peptide Library format

Split the integer strings into fragments with start ('s')
coded by ATG, GTG, TTG and stop codon('*') coded
by TTA, TAG and TGA. Seven consecutive '*' in the
integer coded sequence denotes end of a gene.

The integer represents number
of organisms in which the
heptapeptide is present in the
library. More than 9 occurrence
value is treated as 9.

All possible coding regions (ORFs)

ANN trained on *E. coli-K12*
genome

Predicted protein coding regions

Figure 2: Architecture of Artificial Neural Network

Input-Hidden layer weights

Hidden-Output layer weights

Output layer

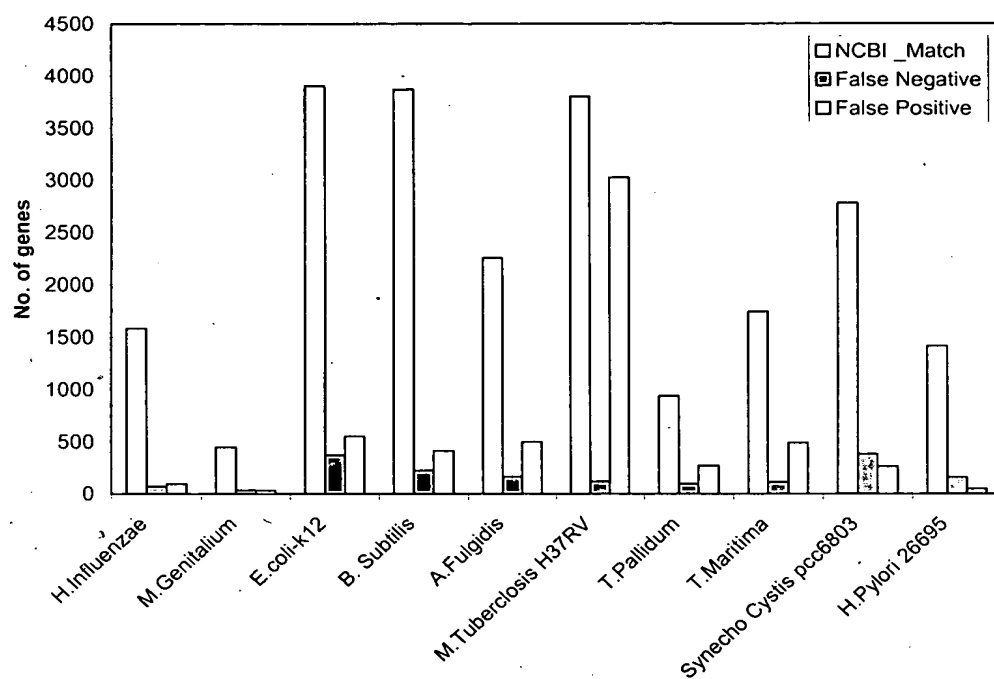Single Output

Input layer
5 Inputs

N = 30

Hidden layer

Fig No. 3 Sensitivity and Specificity of the Method of Invention